These are instructions for using and downloading VariOtator.

Several versions of the VariOtator tool are available:
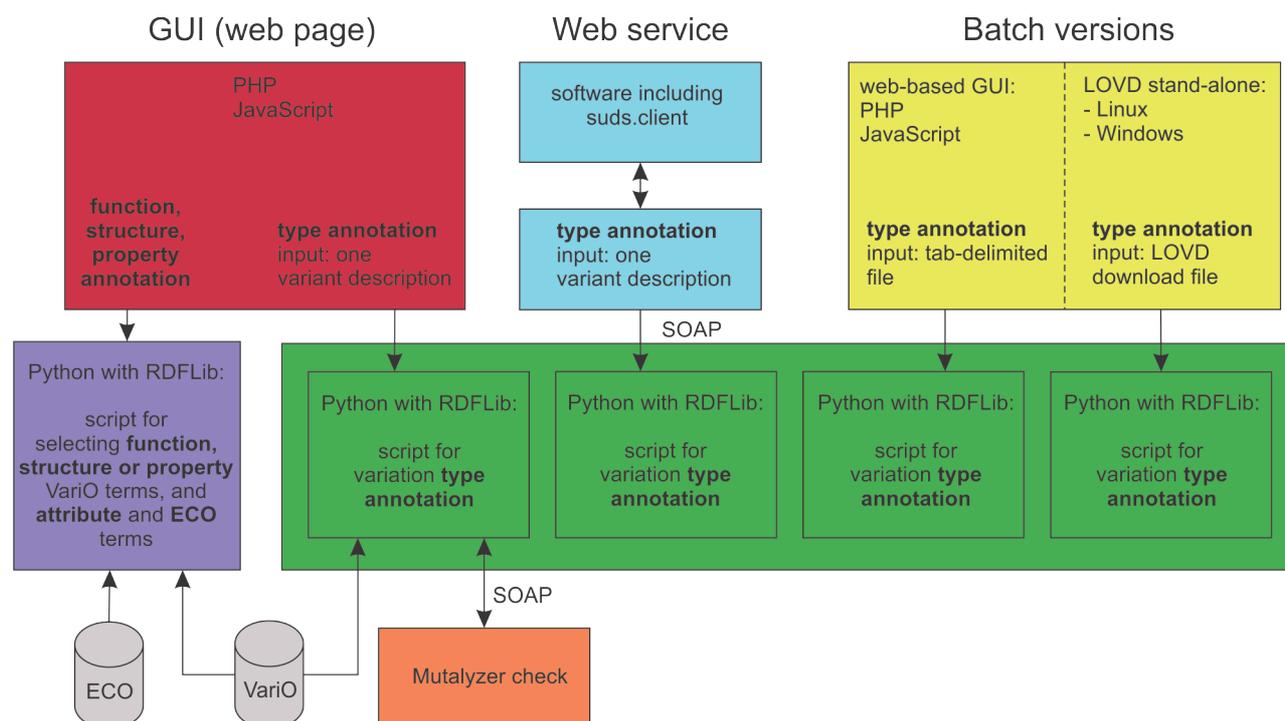

Figure 1: Overview of the VariOtator implementations. The scripts in the green box all have the same functionality, but are optimized for different purposes (mainly regarding to I/O). ECO: Evidence Ontology, format eco.owl; VariO: Variation Ontology, format vario.owl; GUI: Graphical User Interface; LOVD: Leiden Open Variation Database; SOAP: computer network messaging protocol; RDFLib: Python library for working with the Resource Description Framework (RDF).

The web-based Graphical User Interface (GUI; Fig. 1, red box) is for interactive annotation of all kinds of terms. The web service (Fig. 1, light blue box) and the batch versions (Fig. 1, yellow box) are for variation type annotation only.

Use of the GUI:

Type annotation

For variation type annotations, VariOtator accepts a single variant description in Human Genome Variation Society (HGVS; http://www.hgvs.org/mutnomen/) format at a time. It will then generate the VariO annotation automatically. Several input formats are accepted.

Examples of input are:
LRG_1:c.72G>A
NM_003990.3:c.412A>G
NG_008680.1(PAX2):c.412A>G
chr15:g.40702997G>A
r.(21a>u)
p.Trp26Cys
p.W26C

The full lineage of the final VariO term can be obtained by ticking the appropriate box. This means that all parent terms up to (but not including) the root term 'Vario:0001 variation' are included.

Examples of output:

VariO terms for variation type without full lineage:
p.W26C
VariO:0021 amino acid substitution

VariO terms for variation type with full lineage:
p.W26C
VariO:0002 variation affecting protein
VariO:0012 protein variation type
VariO:0325 protein variation classification
VariO:0021 amino acid substitution

When a <u>full coding DNA description</u>, i.e. including the reference sequence (see the first three examples above), is provided, this description is first checked with Mutalyzer (https://mutalyzer.nl) and (predicted) RNA and protein descriptions with their VariO annotations are generated, as well.

Example of output for a full coding DNA description (LRG_128:c.615G>T), including full lineage:

Your variant description was checked with Mutalyzer which provided the following information:

Coding DNA description: LRG_128t1:c.615G>T
Affected protein(s): LRG_128p1:p.(Glu205Asp)

VariO terms for variation type:

c.615G>T
VariO:0128 variation affecting DNA
VariO:0129 DNA variation type
VariO:0322 DNA variation classification
VariO:0135 DNA chain variation
VariO:0136 DNA substitution
VariO:0316 transversion
r.(615g>u)
VariO:0297 variation affecting RNA
VariO:0306 RNA variation type
VariO:0328 RNA variation classification
VariO:0312 RNA substitution
VariO:0316 transversion
VariO:0308 missense variation
p.(Glu205Asp)
VariO:0002 variation affecting protein
VariO:0012 protein variation type
VariO:0325 protein variation classification
VariO:0021 amino acid substitution

If you want to add more annotation to the present results, just choose one of the levels type, function, structure or property, and optionally full lineage, again, then the new annotation will be added. If you choose 'Restart', you will start with a 'fresh' screen and loose your present results.

The button "Create file with VariO annotation" creates a text file with your present results. You can download this file by clicking the link "click here to download your file" under this button.

<u>Function annotation</u>

For annotation of variation affecting <u>function</u> choose the molecular level (DNA, RNA, protein) first. VariOtator then suggests VariO terms of which you will have to choose one.

Again you can have the full lineage of the final VariO term included by ticking the appropriate box.

And as with type annotation, to add more annotation to the present results, choose one of the levels type, function, structure or property, and optionally full lineage, again, then the new annotation is added. If "Restart" is chosen, a 'fresh' screen will appear and the previous results are lost.

The button "Create file with VariO annotation" creates a text file with your present results available for download.

<u>Structure and property annotation</u>

When using the structure and property annotation you are asked to add attribute and/or Evidence Code (ECO) terms to the VariO annotation. Attribute terms are used as modifiers for the previously chosen term, e.g. by adding "VariO:0290 decreased" to the VariO term for variation affecting property: "VariO:0035 effect on protein solubility". ECO terms are used for describing the methods used to obtain the annotated results.

You can decide not to add attribute or ECO terms by clicking the "I do not want to add ..." button.
By clicking "I want to stop here", you will not go further into the ontology (not choosing child terms), and the last chosen term will be the final annotation term.

Use of the batch versions:

<u>The web-based version accessible through a GUI (figure, yellow box, left side)</u>:

Provide the name of a tab-delimited file containing one coding DNA variant description per line, either with or without reference sequence, e.g. "*NM_003990.3:c.412A>G*" or "*c.412A>G*".
The second column should contain protein descriptions and is optional, it should be separated from the first column with a tab.
The first line of the file should contain the column header(s): "codingDNA" and optionally "protein" (without the quotes).

Enter your email address if you want to be notified when and where the resulting file with the VariO annotations can be downloaded. This tab-delimited file contains the columns "codingDNA", "protein", "predicted_RNA_variation", "DNA_annotation", "RNA_annotation", and "protein_annotation".
The columns "protein" and "protein_annotation" have values only when a protein description was available in the uploaded file.
The columns "predicted_RNA_variation" and "RNA_annotation" will only have values when a predicted RNA description is provided (e.g. a prediction for RNA from c.142-2A>G will not be given).

<u>Stand-alone versions for use with an LOVD download file</u>

There are two stand-alone versions, one for Linux systems and one for Windows.

Installing the Windows version

The Windows version is downloaded by clicking the "Windows" link, and then clicking the filename (LOVD_VariOtator-1.2-win32.msi), which installs the program in the "Program Files" (Windows7) or "Program Files (x86)" folder (Windows 8.1) under the directory/folder "LOVD_VariOtator". After installing there is an executable "addVariOtoLOVD.exe". It is run by double-clicking the filename.

Installing the Linux version

The Linux version is downloaded by clicking the "Linux" link. Unpack the "LOVD_VariOtator_build_1.2.tgz" file. The program and accompanying files are then installed in the directory "build", in the sub-directory "exe.linux-x86_64-2.7". The executable "addVariOtoLOVD" is run by the command "./addVariOtoLOVD".

Running the stand-alone versions:

Before the stand-alone versions can be used, make sure the appropriate VariO annotations columns were added to the LOVD database. You are referred to the LOVD documentation (http://www.lovd.nl/2.0/docs/ or http://www.lovd.nl/3.0/docs/) as how to do this. These columns have to exist in the LOVD download file as they are filled in by the VariOtator program. It is suggested to name these "Variant/VariO/DNA", "Variant/VariO/RNA" and "Variant/VariO/protein".

The Windows version is run by double-clicking the filename "addVariOtoLOVD.exe". This will open the Windows Command Prompt window.
The Linux version is run by opening a terminal and giving the command "./addVariOtoLOVD".

After starting the program, you are asked to provide the name of the LOVD download file, either from LOVD version 2

or version 3. The program recognizes the version. Include the full path, e.g. "\Users\User\Downloads\LOVD2.txt" under Windows, where "User" has to be replaced with your own user name.
You are then asked whether you want to overwrite existing VariO terms or not. If you choose "no", new VariO terms will only be added to empty VariO annotation columns, existing values will be left as they are.
Then enter the names of the columns, first the DNA VariO annotation column, then for RNA and then for protein.

Once the program has finished a message is sent with the URL from where to download the annotated file, along with a log file with warnings, errors and such.
End the program by pressing Enter/Return.

The annotated file is now ready for uploading to LOVD. This can be done by clicking the "Configuration" tab in the LOVD database and choosing "Import data using the LOVD import format (….)" in the "Download & Import" section. Detailed instructions can be found in the LOVD documentation.

Webservice

For programmatic access a SOAP web service was developed, with which VariO type annotation generation can be fully integrated into other programs.
The Web Service Definition Language (WSDL) description of the SOAP webservice (Fig. 1, blue box) is available at URL http://variationontology.org/VariOService/?wsdl.
An example of a Python client script can be found at http://variationontology.org/VariO-client-suds.py.